



## **EarthCube Research Coordination Network**

### **"What About Model Data?" - Best Practices for Preservation and Replicability**

The [RCN project](#) would like to invite you to participate in our third workshop. The goal of this workshop is to tackle open issues in preservation, sharing and curation of data and software from research that utilizes simulations. Some questions we plan to tackle in the upcoming workshop:

- How can simulation data and software curation activities be supported in a sustainable way?
- What is the appropriate archive lifetime of simulation data (e.g., at what point does simulation data lose research value such that ongoing data curation costs are no longer justifiable)?
- How can equity in provenance, accessibility, and dissemination of data and software be ensured?
- How should requirements from publishers and funding agencies play a role in data and software curation? (And what support should they/can they provide?)

When and Where:

**25-27 July, 2022**

**Grand Forks, North Dakota**

\*Participant travel costs will be covered\* We welcome model researchers, data curation specialists, representatives from agencies and publishers, and students interested in these issues to join us in this workshop. If you would like to apply to participate in this workshop, please fill out this [Google Form](#) by Mar. 31.

Project Leaders:

Gretchen Mullendore

Doug Schuster

Matt Mayernik

Jared Marquis

## 1. Project Overview

There is strong agreement across the sciences that replicable workflows are needed for computational modeling. However, recent efforts to standardize data sharing and preservation guidelines within research institutions, professional societies, and academic publishers make clear that the scientific community does not know what to do about data produced as output from computational models. The massive size of the simulation outputs, as well as the large computational cost to produce these outputs, makes this not only a problem of replicability, but also a “big data” problem. While most researchers that produce simulation output would welcome more use of their output products, and many end users would welcome more data availability, the reality is that we are producing far more simulation output than can be reasonably stored in repositories. The goal of this [Research Coordination Network \(RCN\) project](#) is to develop guidance on what data and software elements of simulation based research need to be preserved and shared to meet community open science expectations, including those of funders and publishers. .

## 2. Data Preservation Rubric

The first product of this RCN project is a [rubric](#) to be used to assist a researcher in determining what simulation outputs should be deposited in a FAIR aligned community repository to communicate knowledge. The rubric is built to help researchers make decisions about what simulation output needs to be shared via a repository, i.e. made accessible and preserved a sufficient time to satisfy the requirements of publishers and funding agencies. The rubric is a list of simulation/experiment descriptors, organized into themes. The rubric guides researchers through these descriptors to develop a recommendation on whether the researcher should preserve few simulation outputs, some outputs, or the majority of of the simulation outputs

Although questions about what outputs should be preserved are often broadly dictated by the research proposal and the expected user community, the rubric and use case examples are useful for making informed decisions on what simulation workflow outputs should be preserved and shared (e.g., only selected post-processed outputs), and for communicating with end users about why not all simulation outputs fall into this category.

## 3. For More Information

To find a current list of the project’s outputs, including [data archiving rubric and guidance documents](#), please visit the project website: <https://modeldatarcn.github.io/>.